SUPER AI ENGINEER SEASON 6:

Individual Hackathon & Challenge Hackathon

6 JUN 2026

Schedule

- 9:00 - 10:00 Assessment Testing
- **15** Mins Break time
- 10:15 - 15:10 Individual Hackathon

Internet Connection

ฝั่งที่นั่ง

STAGE

A1	B1	C1	D1	E1	F1	G1	H1	I1	J1	K1	L1
A2	B2	C2	D2	E2	F2	G2	H2	I2	J2	K2	L2
A3	B3	C3	D3	E3	F3	G3	H3	I3	J3	K3	L3
A4	B4	C4	D4	E4	F4	G4	H4	I4	J4	K4	L4
A5	B5	C5	D5	E5	F5	G5	H5	I5	J5	K5	L5
A6	B6	C6	D6	E6	F6	G6	H6	I6	J6	K6	L6
A7	B7	C7	D7	E7	F7	G7	H7	I7	J7	K7	L7
A8	B8	C8	D8	E8	F8	G8	H8	I8	J8	K8	L8
A9	B9	C9	D9	E9	F9	G9	H9	I9	J9	K9	L9
A10	B10	C10	D10	E10	F10	G10	H10	I10	J10	K10	L10
A11	B11	C11	D11	E11	F11	G11	H11	I11	J11	K11	L11
A12	B12	C12	D12	E12	F12	G12	H12	I12	J12	K12	L12
A13	B13	C13	D13	E13	F13	G13	H13	I13	J13	K13	L13
A14	B14	C14	D14	E14	F14	G14	H14	I14	J14	K14	L14

ENTRANCE

Rounter Connection

SuperAISS6_1
SuperAISS6_2
SuperAISS6_3
SuperAISS6_4
SuperAISS6_5

Schedule

9:00 - 10:00 Assessment Testing	Break time 15 min	10:15 - 15:10 Individual Hackathon

Section 1:

Assessmest Testing

9:00 - 10:00

ทดสอบระบบ (Test)

ขั้นตอนทดลองระบบก่อนเวลาสอบจริง

- login เข้าระบบ OAP
- ทำข้อสอบ "ทดลองระบบสอบ (Level 2)"

โดยให้กดเริ่มทำตามแถวที่นั่งสอบ **ครั้งละ 1 แถว** เพื่อไม่ให้ load ช้าในช่วงดึงข้อสอบ

รายละเอียดข้อสอบ

- ข้อสอบ "Academic Examination Super AI Engineer Season 6 (Level 2)"
- เริ่มทำได้ตั้งแต่ วันเสาร์ที่ 6/6 เวลา 09:00-10:00น (แบบ on-demand / 1 ครั้ง)
- เวลาสอบ 90 นาที
- คะแนนเต็ม 100 - ไม่แสดงคะแนนหลังสอบ
- มี Test set 2 ชุด ที่แบ่งเอาไว้ตั้งแต่แรก

ในการสอบจริง วันเสาร์ที่ 6/6
ให้ login เข้าระบบ OAP
โดยให้กดเริ่มทำตามแถวที่นั่งสอบ ครั้งละ 1 แถว เพื่อไม่ให้ load ช้าในช่วงดึงข้อสอบ

Section 2:

Individual Hackathon

5 Domain Hackathon

10:15 - 15:10

5 Domain Hackathon

- Thai Call Center ASR
- Chest Disease Detection
- Thai Math VQA
- Sleep Stage Classification
- Heart Disease

ไม่อนุญาตให้ใช้ API ในการส่ง Submission

Thai Call Center ASR

Natural Language Processing

What's in the Audio?

Complex Banking Queries

คำถามด้านธนาคารที่ซับซ้อน

- อายัด / ยกเลิกบัตรเครดิต
- สอบถามสินเชื่อ (loan inquiries)
- แก้ปัญหาบัญชี / โปรโมชั่น

Conversational Nuances

ความเป็นธรรมชาติของบทสนทนา

- Backchannels: "อืม", "ค่ะ", "ครับ"
- Pauses & hesitations กลางประโยค
- พูดทับ / ขัดจังหวะ (interruptions)

Extreme Audio Conditions

สภาพเสียงที่ท้าทายโมเดล

- Phone-line compression (8 kHz)
- Gaussian background noise
- Pitch shift & speed (fast/slow)

How You're Scored

ERROR RATE

$$\frac{S + D + I}{N}$$

- **S** Substitutions — ตัวอักษรที่ถูกแทน
- **D** Deletions — ตัวอักษรที่หายไป
- **I** Insertions — ตัวอักษรที่เกินมา
- **N** Total chars/words in ground truth

SUBMISSION FORMAT

CSV with header — file_name, text

```
file_name,text
RSP_001_audio.wav,ผมทำบัตรเครดิตหายครับ...
BCH_002_audio_noise.wav,สวัสดีค่ะพอดี...
SDB_015_turn_1_fast.wav,มีโปรโมชั่นอะไร...
```

Lower error rate = Better. ระวัง *special chars* ที่ไม่ได้พูดจริง

Thai Math VQA

Natural Language Processing

Dataset description

⚠️ **Total :**

↳ **700 problem images**

⚠️ **Split :**

↳ **Train 280 / Test 420**

⚠️ **Per item :**

↳ **1 image → 1 answer**

9 source buckets

Bucket	Count
101	26
102	27
103	125
104	134
105	101
116	105
118	38
120	30
122	114

Task & Answer formats

- **Input :** ภาพ JPG 1 ภาพ — มีข้อความไทย + ไดอะแกรม + สัญกรณ์คณิตศาสตร์
- **Output :** string คำตอบสั้น 1 ค่า (ส่งเป็น CSV: id, answer)
- **Topics :** arithmetic · algebra · geometry · combinatorics · word problems
- **Grades :** primary → lower-secondary → upper-secondary
- **Answer can be integers, decimals, LaTeX, Thai-unit phrases, short Thai text**

Answer format examples

รูปแบบ	ตัวอย่าง
จำนวนเต็ม / ทศนิยม	36 · -2 · 2.8125
เลข + หน่วยภาษาไทย	20 ตารางเซนติเมตร · 30 องศา
LaTeX	$6\sqrt{3}$ · $\frac{17}{10}$
วลีไทยสั้น	ขาดทุน ร้อยละ 1 · ข้อ (จ)

Evaluation

Evaluation Metric: Exact-match Accuracy (after normalization)

- **Public Leaderboard: ≈ 34% ของ test set**
- **Private Leaderboard: ≈ 66% (ใช้ตัดสินจริง)**

⚠️ **Normalize pipeline :**

- lowercase + strip whitespace
- Thai digits ๐–๙ → 0–9
- ตัด $ และหน่วย (องศา, หน่วย, บาท …)
- expand LaTeX: \frac, \sqrt, \pi
- ตัด { } \ , + canonical integer (2.0 → 2)

ข้อห้ามสำคัญ

ห้ามใช้ commercial VLM API

↓

✓ ใช้เฉพาะ open-weights ที่รันเอง

Sleep Stage Classification

Signal Processing

Example dataset

Features

labels

t_0

16Hz

t_n

BVP	ACC_X	ACC_Y	ACC_Z	TEMP	EDA	HR	IBI	Sleep_Stage
25.61	-22	-61	5	32.09	0.06532	72.85	1.0625	W
23.45	-22	-61	5	32.09	0.06532	72.85	1.0625	W
21.99	-22	-62	4	32.09	0.06532	72.85	1.0625	W
21.14	-22	-62	4	32.09	0.06532	72.85	1.0625	W
20.52	-19	-61	10	32.09	0.06532	72.85	1.0625	W
19.86	-19	-61	10	32.09	0.06532	72.85	1.0625	W
18.95	-19	-61	6	32.09	0.06532	72.85	1.0625	W
17.75	-19	-61	6	32.09	0.06532	72.85	1.0625	W
16.37	-23	-62	4	32.09	0.06532	72.85	1.0625	W
14.96	-23	-62	4	32.09	0.06532	72.85	1.0625	W
13.44	-24	-63	1	32.09	0.06532	72.85	1.0625	W
11.68	-24	-63	1	32.09	0.06532	72.85	1.0625	W
9.59	-21	-63	5	32.09	0.06532	72.85	1.0625	W
7.09	-21	-63	5	32.09	0.06532	72.85	1.0625	W
4.27	-19	-61	9	32.09	0.06532	72.85	1.0625	W
1.48	-19	-61	9	32.09	0.06532	72.85	1.0625	W

Features

- **BVP :** Blood volume pulse derived from PPG sensor
- **IBI (ms) :** Inter-beat interval (between individual heart beats)
- **EDA (µS) :** Electrodermal activity from the galvanic skin response sensor
- **TEMP (°C) :** Skin temperature from the infrared
- **ACC :** Triaxial accelerometry with each axis named ACC_X, ACC_Y, ACC_Z
- **HR (bpm) :** Heart rate, estimated from the BVP
- **Sleep Stages :** The technician-annotated sleep labels from PSG, recorded every 30 seconds, **The possible values are N1, N2, N3, R, and W.**

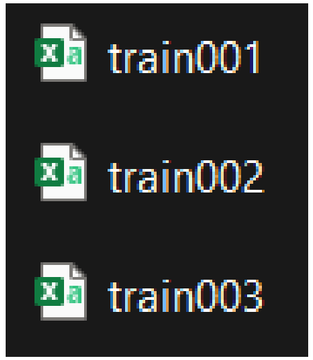
Dataset description

❗Sampling rate: 16Hz

↳ 1 second = 16 points

❗Segmentation : 30 seconds

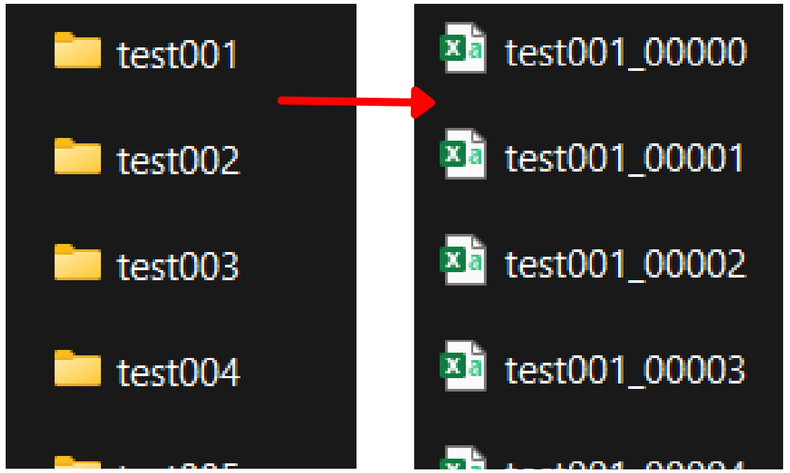
↳ 1 label = 30 seconds



Train set : 83 subjects

Test set : 10 subjects

❗Segments : each 30 s



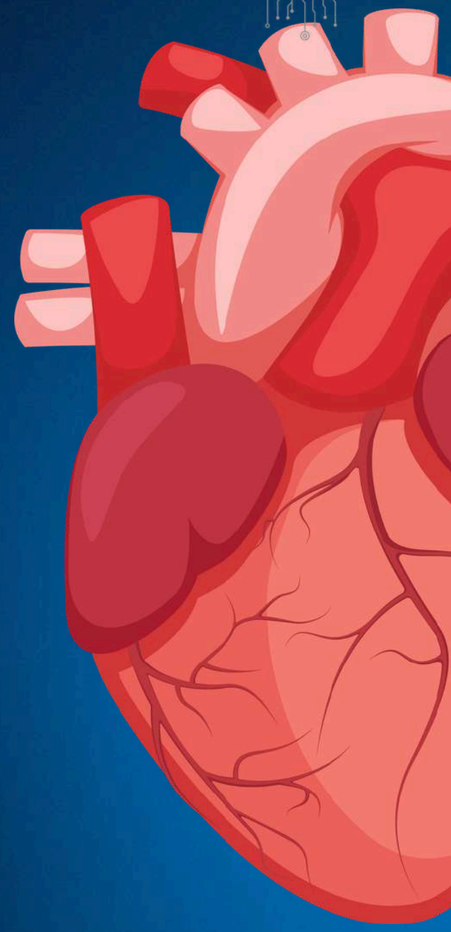
id	labels
test001_00000	W
test001_00001	W
test001_00002	W
test001_00003	W
test001_00004	W

Evaluation

- **Evaluation Metric: F1-Score (weighted)**
  - Public Leaderboard: 50%
  - Private Leaderboard: 50%

Heart Disease Prediction

Data Science



Feature columns

- **History of HeartDisease or Attack**: ประวัติของโรคหัวใจหรือเคยมีอาการ,
- **High Blood Pressure**: ความดันโลหิตสูง,
- **Told High Cholesterol**: เคยได้รับการบอกว่ามีคอเลสเตอรอลสูง,
- **Cholesterol Checked**: เคยตรวจระดับคอเลสเตอรอล,
- **Body Mass Index**: ดัชนีมวลกาย (BMI),
- **Smoked 100+ Cigarettes**: เคยสูบบุหรี่อย่างน้อย 100 มวนในชีวิต,
- **Diagnosed Stroke**: เคยได้รับการวินิจฉัยว่าเป็นโรคหลอดเลือดสมอง (Stroke),
- **Diagnosed Diabetes**: เคยได้รับการวินิจฉัยว่าเป็นโรคเบาหวาน,
- **Leisure Physical Activity**: มีกิจกรรมเพื่อการพักผ่อนหรือออกกำลังกาย,
- Fruit or Vegetable Intake (1+ per Day)': 'บริโภคผลไม้หรือผัก อย่างน้อยวันละ 1 ครั้ง',
- Heavy Alcohol Consumption': 'บริโภคแอลกอฮอล์ในปริมาณมาก',
- Health Care Coverage': มีประกันสุขภาพหรือได้รับความคุ้มครองด้านสุขภาพ,
- Doctor Visit Cost Barrier: มีอุปสรรคด้านค่าใช้จ่ายในการพบแพทย์,
- General Health': สุขภาพโดยรวม (การประเมินตนเองเกี่ยวกับสุขภาพ),
- Difficulty Walking: มีความยากลำบากในการเดินหรือเคลื่อนไหว,

- **Sex**: เพศ,
- **Age**: อายุ,
- **Education Level**: ระดับการศึกษา,
- **Income Level**: ระดับรายได้

Train Dataset

Heart Disease Health Indicators

Train: 223,084 rows

Test: 74,361 rows

	ID	History of HeartDisease or Attack	High Blood Pressure	Told High Cholesterol	Cholesterol Checked	Body Mass Index	Smoked 100+ Cigarettes	Diagnosed Stroke	Diagnosed Diabetes	Leisure Physical Activity	...
0	train_000001	No	Yes	Yes	Yes	40.68	Yes	No	No	No	...
1	train_000002	No	No	No	No	24.36	Yes	No	No	Yes	...
2	train_000003	No	Yes	Yes	Yes	27.33	No	No	No	No	...

Train Dataset

Some rows with missing values

	ID	History of HeartDisease or Attack	High Blood Pressure	Told High Cholesterol	Cholesterol Checked	Body Mass Index	Smoked 100+ Cigarettes	Diagnosed Stroke	Diagnosed Diabetes	Leisure Physical Activity
4	train_000005	NaN	Yes	Yes	Yes	34.56	Yes	No	No	Yes
10	train_000011	No	No	NaN	No	50.64	No	No	No	No
11	train_000012	No	Yes	NaN	No	13.64	Yes	No	No	Yes

Train Dataset

Feature types

- Categorical feature
- Numerical feature
- Ordinal feature

General Health	Difficulty Walking	Sex	Education Level	Income Level	Age	Vegetable or Fruit Intake (1+ per Day)
Very Poor	Yes	Male	Some high school	$15,000 to less than $20,000	98	Yes
Poor	Yes	Male	High school graduate	($10,000 to less than $15,000	53	No
Poor	No	Male	Some high school	($10,000 to less than $15,000	65	Yes

Train Dataset

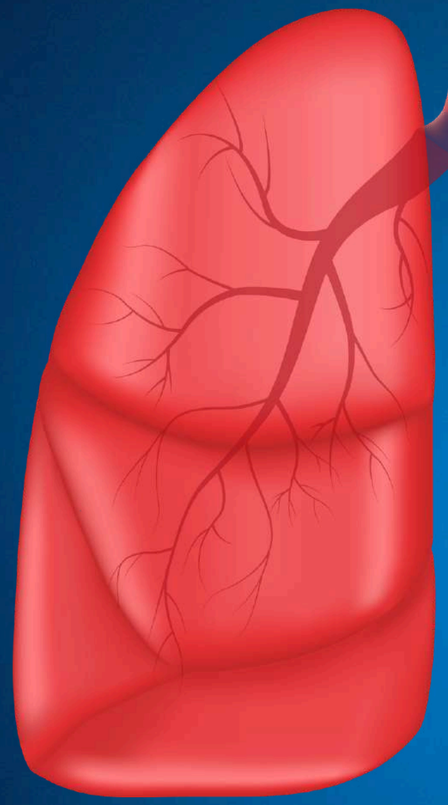
	ID	High Blood Pressure	Told High Cholesterol	Cholesterol Checked	Body Mass Index	Smoked 100+ Cigarettes	Diagnosed Stroke	Diagnosed Diabetes	Leisure Physical Activity	Heavy Alcohol Consumption
0	test_000001	Yes	Yes	Yes	24.84	No	No	No	Yes	No
1	test_000002	Yes	No	Yes	29.08	Yes	No	No	No	No
2	test_000003	Yes	Yes	Yes	35.23	Yes	No	No	No	No
3	test_000004	No	No	Yes	24.78	Yes	No	No	No	No

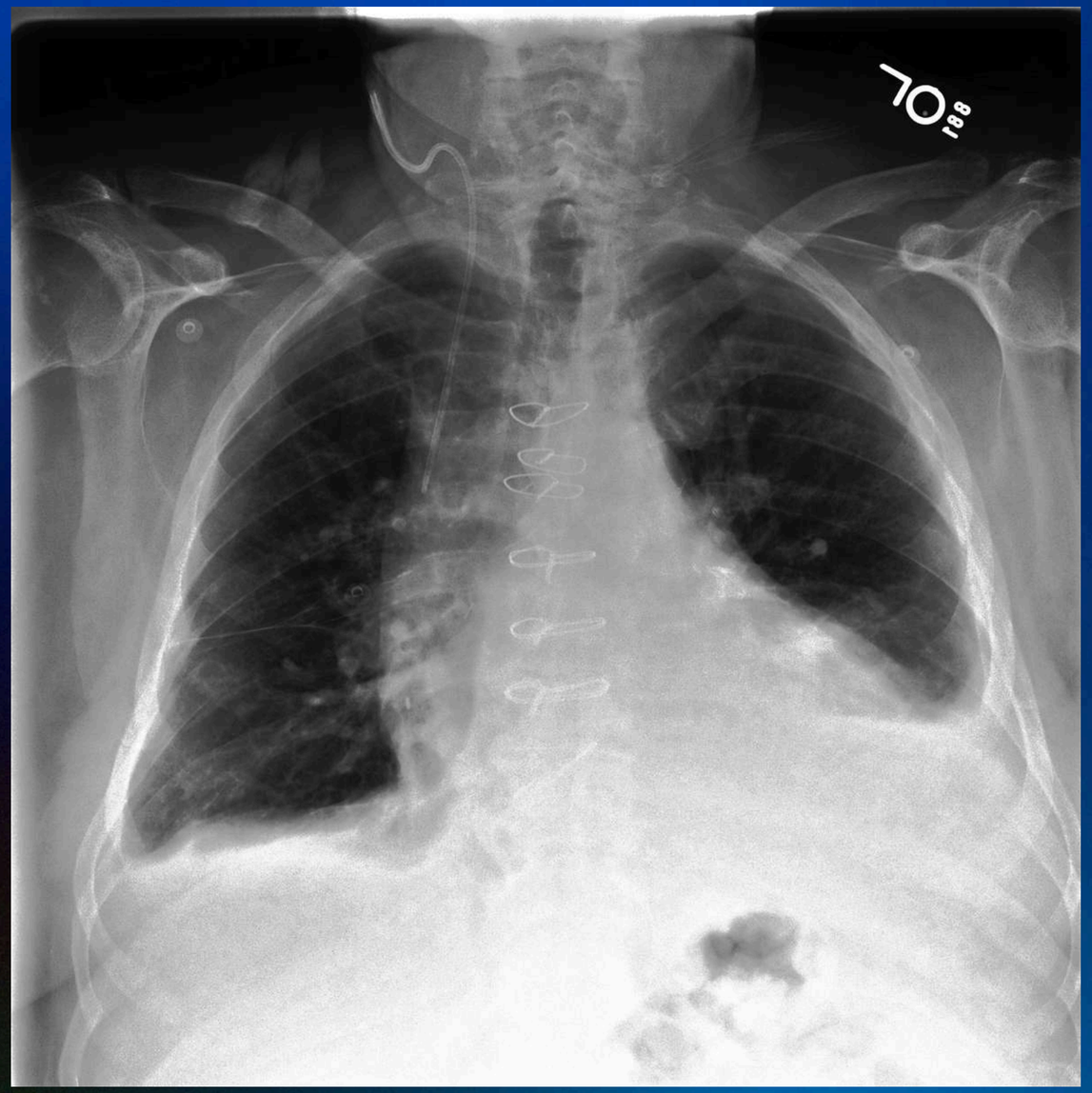
Evaluation

- **Evaluation Metric: F2-score (average)**
  - Public Leaderboard: 50%
  - Private Leaderboard: 50%

Chest Disease Detection

Image





Predict multiple chest X-ray conditions from images across 9 possible classes

- **Atelectasis** : ภาวะปอดแฟบ
- **Cardiomegaly** : โรคหัวใจโต
- **Consolidation** :ปอดแน่น / ปอดอักเสบแบบแน่น
- **Edema** : อาการบวมน้ำในปอด
- **Enlarged Cardio mediastinum** :หัวใจและช่องกลางอกโต
- **Fracture** : กระดูกหัก
- **Lung Lesion** : รอยโรคในปอด
- **Lung Opacity** : ความทึบในปอด
- **No Finding** : ไม่พบความผิดปกติ (หากเป็น 1, อื่นๆ จะเป็น 0)

Data

- **Training Set with 9963 samples**
- **Test set with 2506 samples**

Files

- image/: Training and Test images.
- train.csv: Contains image_id and label for each training image.
- sample_submission.csv: Format for submission with image_id and predicted label.

Data

train.csv (647.78 kB)

Detail Compact Column

10 of 14 columns

filename	Atelectasis	Cardiomegaly	Consolidation	Edema	Enlarged Cardio...
9963 unique values	0 – 1	0 – 1	0 – 1	0 – 1	0 – 1
cxr00030.jpg	1.0	0.0	0.0	0.0	0.0
cxr00033.jpg	0.0	0.0	0.0	0.0	0.0
cxr00034.jpg	0.0	0.0	0.0	0.0	0.0
cxr00035.jpg	0.0	0.0	0.0	0.0	0.0
cxr00038.jpg	0.0	0.0	0.0	0.0	1.0
cxr00042.jpg	0.0	0.0	0.0	0.0	1.0
cxr00044.jpg	0.0	0.0	0.0	0.0	0.0
cxr00047.jpg	0.0	1.0	0.0	0.0	0.0
cxr00048.jpg	0.0	1.0	0.0	1.0	0.0
cxr00049.jpg	0.0	0.0	0.0	1.0	0.0
cxr00051.jpg	0.0	0.0	0.0	1.0	0.0
cxr00052.jpg	0.0	0.0	0.0	0.0	0.0
cxr00057.jpg	0.0	1.0	0.0	1.0	0.0

Data

test_submission.csv (65.45 kB)

Detail Compact Column

10 of 14 columns

About this file

This file does not have a description yet.

filename	Atelectasis	Cardiomegaly	Consolidation	Edema	Enlarged Cardio...
2506 unique values	2506 total values	2506 total values	2506 total values	2506 total values	2506 total values
cxr00001.jpg	0.0	0.0	0.0	0.0	0.0
cxr00002.jpg	0.0	0.0	0.0	0.0	0.0
cxr00003.jpg	0.0	1.0	0.0	0.0	0.0
cxr00007.jpg					
cxr00008.jpg					
cxr00010.jpg					
cxr00011.jpg					

Evaluation

- **Evaluation Metric: sample-average F1**
  - Public Leaderboard: 50%
  - Private Leaderboard: 50%

Rules

- ผู้เข้าอบรมต้องใช้ คอมพิวเตอร์ส่วนตัวในการสอบ
- สามารถใช้เครื่องมือ เช่น VSCode, Google Colab, Kaggle หรือแพลตฟอร์มอื่นๆ
- ผู้เข้าอบรมจะได้รับทรัพยากร LANTA 100 SHr ต่อคน
- สามารถใช้ Generative AI ทุกชนิด (เช่น ChatGPT, GitHub Copilot ฯลฯ) ในการเขียนโค้ด หรือสอบถามข้อมูล
- ห้ามใช้โซเชียลมีเดียแพลตฟอร์ม (เช่น Facebook, Discord ฯลฯ) ตลอดการแข่งขัน
- อนุญาตให้ ค้นหาข้อมูลบนอินเทอร์เน็ต หรือเปิด แหล่งข้อมูลที่เตรียมไว้ล่วงหน้า
- ห้ามแชร์ข้อมูลหรือโค้ดกับผู้อื่นระหว่างการสอบ
- ห้ามติดต่อกับบุคคลภายนอก เว้นแต่ได้รับอนุญาตจากกรรมการ (เช่น ผู้ปกครองในกรณีจำเป็น)
- ห้ามนำอุปกรณ์คอมพิวเตอร์ออกจากห้องสอบตลอดระยะเวลาการสอบ
- อนุญาตให้ใช้ AutoML ในการทดสอบ
- ห้ามใช้ Third-party API ในการทำ Submission ขึ้นมา (เอา LLM เจนคำตอบโดยตรง)
- ห้ามใช้ Private Dataset

Important Information

ระยะเวลาการแข่ง : Sat 6 Jul (10:40 – 15:40)
จำนวนครั้งต่อ hack ในการส่ง : 5 ครั้ง ตัดรอบ 15.40 น. (ต่อ hackathon)

Individual Test: Sleep Stage Classification
https://www.kaggle.com/t/aa4ecc875124416cadc5b24cbc850b9a

Individual Test: Heart Disease Prediction
https://www.kaggle.com/t/01ed6ef789f94da3b1255be4a4c0850f

Individual Test: Thai Call Center ASR
https://www.kaggle.com/t/32a48a849e9642ec99611df67c6a85fc

Individual Test: Math VQA Challenge
https://www.kaggle.com/t/5fb178288a334665bda6e26eb6fb844a

Individual Test: Chest Disease Prediction
https://www.kaggle.com/t/1c5717bb90d04bac9fc04b0bcebd74c3

Large datasets path on LANTA

for

Individual Test: Chest Disease Detection

/project/zz992000-zdevb/zz992000/individual-test-chest-disease-detection

Goodluck